# CONTROL OF GENE EXPRESSION

The present invention relates to the control of gene expression and, in particular, it relates to methods of, and means for, suppressing the expression of a particular, selected gene.

The ability to selectively suppress the expression of a gene is useful in many areas of biology, for example in methods of treatment where the expression of the gene may be undesirable; in preparing models of disease where lack of expression of a particular gene is associated with the disease; in modifying the phenotype in order to produce desirable properties. Thus, the ability to selectively suppress the expression of a gene may allow the "knockout" of human genes in human cells (whether wild type or mutant) and the knockout of eukaryotic genes in studies of development and differentiation.

Present methods of attempting to suppress the expression of a particular gene fall into three main categories, namely antisense technology, ribozyme technology and targeted gene deletion brought about by homologous recombination.

Antisense techniques rely on the introduction of a nucleic acid molecule into a cell which typically is complementary to a mRNA expressed by the selected gene. The antisense molecule typically suppresses translation of the mRNA molecule and prevents the expression of the polypeptide encoded by the gene. Modifications of the antisense technique may prevent the transcription of the selected gene by the antisense molecule binding to the gene's DNA to form a triple helix.

Ribozyme techniques rely on the introduction of a nucleic acid molecule into a cell which expresses a RNA molecule which binds to, and catalyses the selective cleavage of, a target RNA molecule. The target RNA molecule is typically a mRNA molecule, but it may be, for example, a retroviral RNA molecule.

Antisense- and ribozyme-based techniques have proven difficult to implement and they show varying degrees of success in target gene suppression or inactivation. Furthermore, these two techniques require persistent expression or administration of the gene-inactivating agent.

Targeted gene deletion by homologous recombination requires two gene-inactivating events (one for each allele) and is not easily applicable to primary cells, particularly for example primary human mammary cells which can only be maintained in culture for a few passages. Targeted gene deletion has remained difficult to perform in plants. The *cre-lox* mediated site-specific integration has been the method of choice although the efficiency of specific integrative events is low (Alberts *et al* (1995) *Plant J.* 7, 649-659; Vergunst & Hooykass (1998) *Plant Mol. Biol.* **38**, 393-406; Vergunst *et al* (1998) *Nucl. Acids Res.* **26**, 2729-2734).

These major shortcomings in existing technology have led us to seek an alternative strategy.

Acute promyelocytic leukaemia (APL) is underlined by the involvement of mutant retinoic acid receptor (RAR) proteins, formed by gene fusions brought about by chromosomal translocations. Molecular analysis of one

APL subset has identified a fusion between the RARα gene and a Kruppel-like zinc finger gene named promyelocytic leukaemia zinc finger (PLZF). Further investigations have shown that the resulting PLZF-RARα fusion protein functions as a gene repressor by targeting histone deacetylation of retinoic acid regulated genes. Several studies have shown that this repression is mediated by the PLZF portion of the fusion protein, which interacts with a complex of proteins which includes the components N-CoR, SMRT, Sin3 and HDAC and which in turn results in the recruitment of the histone deacetylase (HDAC) complex to target genes (see, for example, Grignani *et al* (1998) *Nature* **391**, 815-818; Chen *et al* (1993) *EMBO J* **12**, 1161-1167; Razin (1998) *EMBO J.* **17**, 4905-4908; David *et al* (1998) *Oncogene* **16**, 2549-2556; and Lin *et al* (1998) *Nature* **391**, 811-814). HDAC directed gene inactivation, therefore results from the targeted assembly of components, some of which have been identified (eg N-CoR, SMRT, Sin3 etc) making a gene inactivating complex which mediates its effect through histone deacetylation.

Although this work shows that in certain forms of APL fusion proteins are able to recruit histone deacetylase activity which appears to have the effect of inactivating the expression of certain genes, no-one has suggested that a method can be devised based on recruitment of histone deacetylase or other means of inactivating chromatin in order to selectively suppress expression of a chosen target gene or a set of genes. Surprisingly, we have shown that this can be achieved.

RARα-PLZF and RARα-PML fusion proteins are known from studies of acute promyelocytic leukaemia (APL) and are described in, for example, Grignani *et al* (1998) *Nature* **391**, 815-818.

Fusions of GAL4 with a portion of PLZF protein, and LexA DNA binding domain fused to various fragment of Sin 3A are described in David *et al* (1998) *Oncogene* **16**, 2549-2556 which, for the avoidance of doubt, are excluded from the polypeptide of the present invention. Fusions of the GAL4 DNA binding domain and PLZF-RARα are described in Lin *et al* (1998) *Nature* **391**, 811-814 which, for the avoidance of doubt, are excluded from the polypeptide of the present invention.

Fusions of the GAL4 DNA binding domain with N-CoR or portions thereof, or with the C terminal domain of the $T_3R\beta1$ receptor molecule (thyroid hormone receptor molecule), and LexA DNA binding domain fused with the C terminal domain of the $T_3R\alpha$ or RARα (retinoic acid receptor) receptor molecules, which, for the avoidance of doubt, are excluded from the polypeptide of the present invention, are described in Hörlein *et al* (1995) *Nature* **377**, 397-404. Fusions of the GAL4 DNA binding domain with the C terminal domain of vErbA (viral oncogene erbA of the avian erythroblastosis virus (AEV)), $T_3R$ and RAR receptor molecules are also mentioned. These polypeptides are also, for the avoidance of doubt, excluded from the polypeptide of the present invention.

There is no suggestion in David *et al* (1998) *Oncogene* **16**, 2549-2556, Lin *et al* (1998) *Nature* **391**, 811-814 or Hörlein *et al* (1995) *Nature* **377**, 397-404 that polypeptides comprising a nucleic acid binding portion and a chromatin inactivation portion can be designed and engineered to bring about the selective suppression of a chosen gene. Rather, David *et al*

(1998) and Lin *et al* (1998) are both studies of gene repression in acute promyelocytic leukaemia, and Hörlein *et al* (1995) relates to the identification of N-CoR.

A first aspect of the invention provides a polypeptide comprising a nucleic acid binding portion which binds to a site present in a eukaryotic genome and a chromatin inactivation portion provided that when the nucleic acid binding portion is a DNA binding portion of RARα the chromatin inactivation portion is not a portion of PLZF protein and is not a portion of PML protein; and provided that when the nucleic acid binding portion is a DNA binding portion of the *Saccharomyces cerevisiae* GAL4 protein the chromatin inactivation portion is not a portion of PLZF protein, the C-terminal domain of vErbA, $T_3R$, $T_3R\beta1$ or RAR, or N-CoR or a portion of N-CoR; and provided that when the nucleic acid binding portion is a DNA binding portion of the *Escherichia coli* LexA the chromatin inactivation portion is not mSin3, or the C-terminal domain of $T_3R\alpha$ or RARα.

The polypeptides of the invention may be useful in methods and uses provided by further aspects of the invention, discussed in more detail below. In particular, the polypeptides of the invention may be useful in a method of suppressing the expression of a selected gene in a eukaryotic cell the method comprising introducing into the cell (a) a polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with the selected gene which site is present in a eukaryotic genome and a chromatin inactivation portion, or (b) a polynucleotide encoding said polypeptide.

It is preferred if the polypeptides of the invention are hybrid polypeptides which do not occur in nature. For example, it is preferred if the nucleic acid binding portion is derived from one protein and that the chromatin inactivation portion is derived from a different protein and that the molecular configuration does not arise in nature, for example through chromosome translocation events. The proteins from which the nucleic acid binding portion and the chromatin inactivation portion are derived may be from the same species (for example, as is described in more detail below, the nucleic acid binding portion may be a DNA binding portion of a human steroid receptor protein such as oestrogen receptor (ER) and the chromatin inactivation portion may be a portion of human PLZF) or they may be from different species (for example a bacterial DNA binding protein may be fused to a portion of human PLZF).

Thus, in a particular preferred embodiment the polypeptide of the invention is one which is produced by genetic engineering means wherein the nucleic acid binding portion and the chromatin inactivation portion are selected as is described in more detail below.

It is preferred if the nucleic acid binding portion is not the *Saccharomyces cerevisiae* GAL4 protein or a DNA-binding portion thereof, and it is preferred if the nucleic acid binding portion is not the *Escherichia coli* LexA protein or a DNA-binding portion thereof.

In relation to the first aspect of the invention the site present in a eukaryotic genome is a site which is at or associated with a selected gene or genes whose expression it is desirable to suppress or inactivate. It is

preferred if the site is a site which is naturally present in a eukaryotic genome. However, as is discussed in more detail below, the site may be one which has been engineered into the genome, or it may be a site associated with an inserted viral sequence. The site engineered into the genome to be in the vicinity of the gene whose expression is to be suppressed may be a site from the same species (but present elsewhere in the genome) or it may be a site present in a different species. By "genome" we include not only chromosomal DNA but other DNA present in the eukaryotic cell, such as DNA which has been introduced into the cell, for example plasmid or viral DNA. It is preferred if the nucleic acid binding portion can bind to chromosomal DNA or, as is described in more detail below, to RNA transcribed from chromosomal DNA.

The chromatin inactivation portion may be any polypeptide or part thereof which directly or indirectly leads to chromatin inactivation. By "directly" leading to chromatin inactivation we mean that the polypeptide or part thereof itself acts on the chromatin to inactivate it. By "indirectly" leading to chromatin inactivation we mean that the polypeptide or part thereof does not itself act on the chromatin but rather it is able to recruit or promote a cellular component to do so.

Chromatin inactivation generally results in the suppression or inactivation of gene expression. Chromatin inactivation is typically a localised event such that suppression or inactivation of gene expression is restricted to, typically, one or a few genes. Thus, the chromatin inactivation portion is any suitable polypeptide which, when part of the polypeptide of the invention and when targeted to a selected gene by the nucleic acid binding portion, locally inactivates the chromatin associated with the selected gene

so that expression of the gene is inactivated or suppressed. Histone deacetylation is associated with chromatin inactivation and so it is particularly preferred if the chromatin inactivation portion facilitates histone deacetylation. Targeted deacetylation of histones associated with a given gene leads to gene inactivation in an, essentially, irreversible manner. By "suppression" or "inactivation" of gene expression we mean that in the presence of the polypeptide of the invention the expression of the selected, targeted gene is at least five-fold, preferably at least ten-fold, more preferably at least 100-fold, and most preferably at least 1000-fold lower than in the absence of the polypeptide of the invention under equivalent conditions. Gene expression can be measured using any suitable method including using reverse transcriptase-polymerase chain reaction (RT-PCR), RNA hybridisation, RNAse protection assays, nuclear run-off assays and alteration of chromatin as judged by DNAse 1 hypersensitivity.

In animal and plant cells histone deacetylation is brought about by the so-called histone deacetylase complex (HDAC) which contains, in addition to one or more histone deacetylase enzymes, ancillary proteins which are involved in the formation and function of the complex. In addition, there are other protein components which although they may not be part of HDAC they bind to or otherwise interact with HDAC and help facilitate histone deacetylation.

Deacetylation and acetylation of histones is a well-known phenomenon which is reviewed in the following: Chen & Li (1998) *Crit. Rev. Eukaryotic Gene Expression* **8**, 169-190; Workman & Kingston (1998) *Ann. Rev. Biochem.* **67**, 545-579; Perlmann & Vennstrom (1995) *Nature*

377, 387- ; Wolfe (1997) *Nature* **387**, 16-17; Grunstein (1997) *Nature* **389**, 349-352; Pazin & Kadonaga (1997) *Cell* **89**, 325-328; DePinho (1998) *Nature* **391**, 533-536; Bestor (1998) *Nature* **393**, 311-312; and Grunstein (1998) *Cell* **93**, 325-328.

The polypeptide composition of the HDAC complex is currently under investigation. Polypeptides which may form part of, or are associated with, certain HDAC complexes include histone deacetylase 1 (HDAC1) Taunton *et al* (1996) *Nature* **272**, 408-441); histone deacetylase 2 (HDAC2) (Yang *et al* (1996) *Proc. Natl. Acad. Sci. USA* **93**, 12845-12850); histone deacetylase 3 (HDAC3) (Dangond *et al* (1998) *Biochem. Biophys. Res. Comm.* **242**, 648-652); N-CoR (Horlein *et al* (1995) *Nature* **377**, 397-404); SMRT (Chen & Evans (1995) *Nature* **377**, 454-457); SAP30 (Zhang *et al* (1998) *Molecular Cell* **1**, 1021-1031). Sin3 (Ayer *et al* (1995) *Cell* **80**, 767-776; Schreiber-Agus *et al* (1995) *Cell* **80**, 777-786) SAP18 (Zhang *et al* (1997) *Cell* **89**, 357-364); and RbAp48 (Qian *et al* (1993) *Nature* **364**, 648-652). All of these papers are incorporated herein by reference. It is believed that there may be further components of the HDAC complex or which interact with the HDAC complex which are, as yet, undiscovered. It is envisaged that these too will be useful in the practice of the invention.

PLZF has been shown to interact with N-CoR and SMRT, which in turn recruit a HDAC complex. PLZF will also directly interact with HDAC (Lin *et al* (1998) *Nature* **391**, 811-814; Grignani *et al* (1998) *Nature* **391**, 815-818; David *et al* (1998) *Oncogene* **16**, 2549-2556).

Complexes formed which contain any of N-CoR, SMRT, Sin3, SAP18, SAP30 and histone deacetylase are described in Heinzel *et al* (1997) *Nature* **387**, 43-48; Alland *et al* (1997) *Nature* **387**, 49-55; Hassig *et al* (1997) *Cell* **89**, 341-347; Laherty *et al* (1997) *Cell* **89**, 349-356; Zhang *et al* (1997) *Cell* **89**, 357-364; Kadosh & Struhl (1997) *Cell* **89**, 365-371; Nagy *et al* (1997) *Cell* **89**, 373-380; and Laherty *et al* (1998) *Molecular Cell* **2**, 33-42. All of these papers are incorporated herein by reference.

Thus, it is particularly preferred if the component of a HDAC complex or the polypeptide which binds to or facilitates recruitment of a HDAC complex is any one of PLZF, SMRT, Sin3, SAP18, SAP30 or N-CoR, or HDACs including HDAC1, HDAC2 or HDAC3. It will be appreciated that it may not be necessary for all of the polypeptides to be present so long as a functional portion thereof is present. For example, with respect to histone deacetylase enzymes (for example, HDAC1, HDAC2 or HDAC3) the functional portion may be a portion that retains histone deacetylase activity or it may be a portion which contains a binding site for other components of a HDAC complex or a portion which otherwise recruits the HDAC complex and promotes histone deacetylation. Similarly, with respect to other components of the HDAC complex or polypeptides which bind to the HDAC complex the functional portion may be a portion which contains a binding site for other components of the HDAC complex. To date six mammalian HDAC genes have been described (Grozinger *et al* (1999) *Proc. Natl. Acad. Sci. USA* **96**, 4868-4873), it is believed that any one or more of these may be useful in the practise of the present invention.

It is preferred that the chromatin inactivation portion is not N-CoR or a portion thereof; or the C-terminal domain of the vErbA, $T_3R$ (including $T_3R\beta1$ or $T_3R\alpha$) or RAR (including RAR$\alpha$) receptor molecule, particularly if the nucleic acid binding portion is the *Saccharomyces cerevisiae* GAL4 protein or a DNA-binding portion thereof, or the *Escherichia coli* LexA protein or a DNA-binding portion thereof.

It is believed that binding motifs are present within the components of the HDAC complex or within polypeptides which bind the HDAC complex and these motifs may be sufficient to act as chromatin inactivation portions in the polypeptide of the invention since they may facilitate histone deacetylation by recruiting a HDAC complex.

Furthermore, it will be appreciated that variants of a component of the HDAC complex or variants of a polypeptide which binds to the HDAC complex may be used. Suitable variants include not only functional portions as described above, but also variants in which amino acid residues have been deleted or replaced or inserted provided that the variant is still able to facilitate histone deacetylation. Thus, suitable variants include variants of histone deacetylase in which the amino acid sequence has been modified compared to wild-type but which retain their ability to deacetylate histones. Similarly, suitable variants include variants of, for example, Sin3 or PLZF in which the amino acid sequence has been modified compared to wild-type but which retain their ability to interact with or in the HDAC complex. Similarly, variants of other proteins interacting with components of the HDAC complex and other transcription factors that can bring about gene inactivation through HDAC activity may be used.

All or parts of the Rb, MAD and MeCpG2 proteins may interact with the HDAC complex.

While most work has been done on HDAC complexes and polypeptides involved in recruiting HDAC complexes in mammalian systems, the fundamental nature of the system is such that functionally equivalent polypeptides are expected to be found in other eukaryotic cells, in particular in other animal cells and in plant cells. For example, Figure 5 shows that polypeptides very closely related to human HDAC1 are present in arabidopsis and in yeast. A plant HDAC complex has been isolated from maize (Lussen *et al* (1997) *Science* **277**, 88-91) and a comparative study of histone deacetylases from plant, fungal and vertebrate cells has been undertaken (Lechner *et al* (1996) *Biochim. Biophys. Acta* **1296**, 181-188). Histone deacetylase inhibitors have been shown to derepress silent rRNA genes in Brassica (Chen & Pickard (1997) *Genes Dev.* **11**, 2124-2136) and a naturally occurring host selective toxin (HC toxin) from *Cochliobolus carbonum* inhibits plant, fungal and mammalian histone deacetylases (Brosch *et al* (1995) *Plant Cell* 7, 1941-1950).

It is not necessary that the chromatin inactivation portion is from the same cell type or species as the cell into which the polypeptide (or polynucleotide encoding the polypeptide) is introduced but it is desirable if it is since such a chromatin inactivation portion may be able to inactivate chromatin more effectively in that cell.

It is particularly preferred if the chromatin inactivation portion of the polypeptide is PLZF, a portion of PLZF that can facilitate histone

deacetylation, or a polypeptide, or portion of a polypeptide, known to cause gene activation *via* histone deacetylation. For example, the portion of PLZF in PLZF-RARα which is involved in APL is believed to interact with N-CoR and SMRT.

It is also particularly preferred if the chromatin inactivation portion is a polypeptide with histone deacetylase enzyme activity such as contained in HDAC1, HDAC2 or HDAC3.

The nucleic acid binding portion may be any suitable binding portion which binds to a site present in a eukaryote, such as a plant or animal, genome. It is particularly preferred that the nucleic acid binding portion is able to bind to a site which is at or associated with a selected gene whose expression is to be suppressed by the presence of the chromatin inactivating portion of the polypeptide of the invention. It is preferred that the nucleic acid binding portion binds selectively to the desired site. There may be one or more desired sites to which the nucleic acid binding portion may bind. For example, the polypeptide of the invention may be used to suppress the expression of a group of genes which each have a binding site for a common DNA binding portion (for example, are under the controls of a steroid hormone receptor such as the oestrogen receptor (ER)). For the avoidance of doubt, the site present in the eukaryote may be a naturally occurring site, or it may be a site which has been engineered to be there. The site need not be originally from the same or any other eukaryote. For example, it may be a bacterial repressor binding site which has been engineered to be present in the DNA of the eukaryotic cell, or it may be a mammalian steroid hormone receptor binding site which has been engineered into plant cells. However, it is preferred if the

site to which the nucleic acid binding portion binds is naturally present in the eukaryotic cell and is present in its natural position in the genome.

The nucleic acid binding portion may be a DNA binding portion or an RNA binding portion. Proteins which have the ability to bind either DNA or RNA in a sequence selective manner are well known in the art and some are described in more detail below. In the case of the RNA binding portion, the site present in the eukaryotic genome which binds the RNA binding portion is, typically, nascent RNA being transcribed from DNA at the selected site for inactivation. The RNA may be that which is being transcribed by the gene whose expression is to be suppressed, or it may be that which is being transcribed by a gene adjacent to, or at least close to, the gene whose expression is to be suppressed. It is preferred that the RNA binding portion binds to an RNA sequence which is at or close to the 5′ end of the transcript. It will be appreciated that whilst being transcribed, nascent RNA remains at or close to its site of transcription and that if the site of transcription is at or close to the gene whose expression is to be suppressed, using an RNA binding portion in the polypeptide of the invention facilitates the localisation of the chromatin inactivation portion to the desired site.

The DNA binding portion may be all or a DNA-binding portion of a zinc-finger DNA binding protein or it may be all or a DNA-binding portion of a helix-turn-helix DNA binding protein.

Suitably the DNA binding portion may be all or a DNA-binding portion of an animal or plant DNA binding protein, or it may be all or a DNA binding portion of a bacterial or yeast DNA binding protein which has

been engineered to bind to one or more sites in the plant or animal genome. Bacterial or yeast DNA binding proteins are less preferred and it is particularly preferred if the DNA binding protein does not contain a DNA binding portion of wild-type *Saccharomyces cerevisiae* GAL4 or wild-type *Escherichia coli* LexA.

Any DNA binding protein with the ability to bind DNA through a DNA recognition sequence may be used. This includes DNA binding proteins, and engineered DNA binding proteins, such as engineered zinc finger proteins and helix-turn-helix DNA binding proteins.

Databases listing transcription factors and their binding sites are listed below:

http://www.embl-heidelberg.de/srs5bin/cgi-bin/wgetz?-fun+pagelibinfo+-info+TFFACTOR
http://www.embl-heidelberg.de/srs5bin/cgi-bin/wgetz?-fun+pagelibinfo+-info+TFSITE
http://www.embl-heidelberg.de/srs5bin/cgi-bin/wgetz?-fun+pagelibinfo+-info+TFCELL
http://www.embl-heidelberg.de/srs5bin/cgi-bin/wgetz?-fun+pagelibinof+-info+TFCLASS
http://www.embl-heidelberg.de/srs5bin/cgi-bin/wgetz?-fun+pagelibinfo+-info+TFMATRIX
http://www.embl-heidelberg.de/srs5bin/cgi-bin/wgetz?-fun+pagelibinfo+-info+TFGENE

It is believed that all or part of the listed transcription factors may be useful in the practice of the invention.

Other gene regulatory proteins which may be useful in the practice of the invention include virally encoded DNA binding proteins such as those required for regulating viral and cellular gene expression and/or viral replication. These include but are not limited to the large T antigen of polyoma viruses, the E2 protein of papillomaviruses and the ICP4, ICP0 protein of herpesviruses.

Sequence specific RNA binding proteins, which bind to nascent RNA, may be engineered to bring about gene inactivation by the methods of the invention through HDAC complex formation in the proximity of transcriptionally active chromatin. For example, the transcriptionally active chromatin may be proviral and the RNA binding protein one which binds to transcribed proviral RNA. The *tat* protein of HIV is an example of an RNA binding protein.

In plants, DNA binding proteins are involved in, amongst other things, floral development, cold regulation/adaptation, and plant responses to ethylene or pathogens. Thus, the polypeptides of the invention, and the methods of the invention described below, may be used to analyse the role of these genes in these developmental and other processes.

A particularly preferred embodiment is wherein the DNA binding portion is all or a DNA binding part of a nuclear receptor DNA binding protein such as a steroid hormone receptor protein.

The nuclear receptor DNA binding protein superfamily includes oestrogen receptor (ER), androgen receptor (AR), progesterone receptor (PR), retinoic acid receptor (RAR) and the like (see Mangelsdorf *et al* (1995) *Cell* **83**, 835-839 for a review and nomenclature).

It is particularly preferred if the steroid hormone receptor protein is estrogen receptor (ER).

As noted, DNA binding proteins may be engineered so as to bind to a particular, selected target DNA sequence which is at or associated with a selected gene. In one embodiment of the invention the DNA binding protein is one which has been engineered to bind to a site which is present in a mutant gene sequence within the plant or animal cell but is not present in the equivalent wild type sequence. For example, and as is discussed in more detail below, the engineered DNA binding portion may bind selectively to a dominant negative, mutated gene, such as a mutant oncogene and, upon binding, chromatin inactivation occurs and suppresses the expression of the mutant oncogene. Examples of oncogenes mutated in human cancer include RAS (*H-ras*) and *Bcl-10*.

Typically, the DNA binding portion and the chromatin inactivation portion are fused such that the fusion polypeptide may be encoded by a nucleic acid molecule. Suitably, the DNA binding portion and the chromatin inactivation portion are joined so that both portions retain their respective activities such that the polypeptide may bind to a site present in a plant or animal genome and, upon binding, the chromatin inactivation portion is still able to inactivate chromatin. The two portions may be joined directly, but they may be joined by a linker peptide. Suitable linker

peptides are those that typically adopt a random coil conformation, for example the polypeptide may contain alanine or proline or a mixture of alanine plus proline residues. Preferably, the linker contains between 10 and 100 amino acid residues, more preferably between 10 and 50 and still more preferably between 10 and 20. In any event, whether or not there is a linker between the portions of the polypeptide the polypeptide is able to bind its target DNA and is able to inactivate chromatin thereby selectively suppressing or inactivating gene expression.

A further aspect of the invention provides a polynucleotide encoding a polypeptide of the invention. In particular, the invention provides a polynucleotide wherein the nucleic acid binding portion and the chromatin inactivation portion are fused such that the fusion polypeptide is encoded by a single open reading frame of the polynucleotide. The polynucleotide may be DNA or RNA; DNA is preferred. DNA may or may not contain introns but, in any case, the polynucleotide encodes a polypeptide of the invention.

Polynucleotides which encode suitable nucleic acid binding portions, particularly DNA binding portions are known in the art or can be readily designed from known sequences such as from known sequences contained in scientific publications or contained in nucleotide sequence databases such as the GenBank, EMBL and dbEST databases. Polynucleotides which encode suitable chromatin inactivation portions are known in the art or can readily be designed from known sequences and made. Polynucleotide sequences encoding various suitable chromatin inactivation portions are given above in the references which refer to the polypeptides

or are available from GenBank or EMBL or dbEST. A reference for PLZF is Chen *et al* (1993) *EMBO J.* **12**, 1161-1167.

Polynucleotides which encode suitable linker peptides can readily be designed from linker peptide sequences and made.

Thus, polynucleotides which encode the polypeptides of the invention can readily be constructed using well known genetic engineering techniques.

A variety of methods have been developed to operably link polynucleotides, especially DNA, to other polynucleotides, including vectors, for example *via* complementary cohesive termini. For instance, complementary homopolymer tracts can be added to the DNA segment to be inserted to the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules.

Synthetic linkers containing one or more restriction sites provide an alternative method of joining the DNA segment to vectors. The DNA segment, generated by endonuclease restriction digestion as described earlier, is treated with bacteriophage T4 DNA polymerase or *E. coli* DNA polymerase I, enzymes that remove protruding, 3'-single-stranded termini with their 3'-5'-exonucleolytic activities, and fill in recessed 3'-ends with their polymerising activities.

The combination of these activities therefore generates blunt-ended DNA segments. The blunt-ended segments are then incubated with a large molar excess of linker molecules in the presence of an enzyme that is able

to catalyse the ligation of blunt-ended DNA molecules, such as bacteriophage T4 DNA ligase. Thus, the products of the reaction are DNA segments carrying polymeric linker sequences at their ends. These DNA segments are then cleaved with the appropriate restriction enzyme and ligated to an expression vector that has been cleaved with an enzyme that produces termini compatible with those of the DNA segment.

Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including International Biotechnologies Inc, New Haven, CN, USA.

A desirable way to modify the DNA encoding the polypeptide of the invention is to use the polymerase chain reaction as disclosed by Saiki *et al* (1988) *Science* **239**, 487-491. This method may be used for introducing the DNA into a suitable vector, for example by engineering in suitable restriction sites, or it may be used to modify the DNA in other useful ways as is known in the art.

In this method the DNA to be enzymatically amplified is flanked by two specific primers which themselves become incorporated into the amplified DNA. The said specific primers may contain restriction endonuclease recognition sites which can be used for cloning into expression vectors using methods known in the art.

Methods of joining a polynucleotide to a nucleic acid vector are, of course, applicable to joining any polynucleotides.

The DNA (or in the case of retroviral vectors, RNA) is then expressed in

a suitable host to produce a polypeptide of the invention. Thus, the DNA encoding the polypeptide of the invention may be used in accordance with known techniques, appropriately modified in view of the teachings contained herein, to construct an expression vector, which is then used to transform an appropriate host cell for the expression and production of the polypeptide of the invention. Such techniques include those disclosed in US Patent Nos. 4,440,859 issued 3 April 1984 to Rutter *et al*, 4,530,901 issued 23 July 1985 to Weissman, 4,582,800 issued 15 April 1986 to Crowl, 4,677,063 issued 30 June 1987 to Mark *et al*, 4,678,751 issued 7 July 1987 to Goeddel, 4,704,362 issued 3 November 1987 to Itakura *et al*, 4,710,463 issued 1 December 1987 to Murray, 4,757,006 issued 12 July 1988 to Toole, Jr. *et al*, 4,766,075 issued 23 August 1988 to Goeddel *et al* and 4,810,648 issued 7 March 1989 to Stalker, all of which are incorporated herein by reference.

The DNA (or in the case of retroviral vectors, RNA) encoding the polypeptide of the invention may be joined to a wide variety of other DNA sequences for introduction into an appropriate host. The companion DNA will depend upon the nature of the host, the manner of the introduction of the DNA into the host, and whether episomal maintenance or integration is desired.

Generally, the DNA is inserted into an expression vector, such as a plasmid, in proper orientation and correct reading frame for expression. If necessary, the DNA may be linked to the appropriate transcriptional and translational regulatory control nucleotide sequences recognised by the desired host, although such controls are generally available in the expression vector. The vector is then introduced into the host through

standard techniques. Generally, not all of the hosts will be transformed by the vector. Therefore, it will be necessary to select for transformed host cells. One selection technique involves incorporating into the expression vector a DNA sequence, with any necessary control elements, that codes for a selectable trait in the transformed cell, such as antibiotic resistance. Alternatively, the gene for such selectable trait can be on another vector, which is used to co-transform the desired host cell.

Host cells that have been transformed by the recombinant DNA of the invention are then cultured for a sufficient time and under appropriate conditions known to those skilled in the art in view of the teachings disclosed herein to permit the expression of the polypeptide, which can then be recovered.

Many expression systems are known, including bacteria (for example *E. coli* and *Bacillus subtilis*), yeasts (for example *Saccharomyces cerevisiae*), filamentous fungi (for example *Aspergillus*), plant cells, animal cells and insect cells.

The vectors include a prokaryotic replicon, such as the ColE1 *ori*, for propagation in a prokaryote, even if the vector is to be used for expression in other, non-prokaryotic, cell types. The vectors can also include an appropriate promoter such as a prokaryotic promoter capable of directing the expression (transcription and translation) of the genes in a bacterial host cell, such as *E. coli*, transformed therewith.

A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur.

Promoter sequences compatible with exemplary bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA segment of the present invention. It is preferred that the promoter is one which can be regulated. It is particularly preferred if the promoter is an inducible promoter which can be selectively induced at an appropriate time once the vector has been introduced into the eukaryotic cell. It will be appreciated that upon induction, the polypeptide of the invention may be expressed in the cell and exert its effect. In this situation, induction of expression of the polypeptide of the invention leads to suppression of the targeted gene. Inducible promoters are known in the art for many eukaryotic cells including plant and animal cells. These include heat-shock-, glucocorticoid-, oestradiol-, and metal-inducible promoter systems.

Typical prokaryotic vector plasmids are pUC18, pUC19, pBR322 and pBR329 available from Biorad Laboratories, (Richmond, CA, USA) and p*Trc*99A and pKK223-3 available from Pharmacia, Piscataway, NJ, USA.

A typical mammalian cell vector plasmid is pSVL available from Pharmacia, Piscataway, NJ, USA. This vector uses the SV40 late promoter to drive expression of cloned genes, the highest level of expression being found in T antigen-producing cells, such as COS-1 cells.

An example of an inducible mammalian expression vector is pMSG, also available from Pharmacia. This vector uses the glucocorticoid-inducible promoter of the mouse mammary tumour virus long terminal repeat to drive expression of the cloned gene.

Useful yeast plasmid vectors are pRS403-406 and pRS413-416 and are generally available from Stratagene Cloning Systems, La Jolla, CA 92037, USA. Plasmids pRS403, pRS404, pRS405 and pRS406 are Yeast Integrating plasmids (YIps) and incorporate the yeast selectable markers *HIS3, TRP1, LEU2* and *URA3*. Plasmids pRS413-416 are Yeast Centromere plasmids (YCps).

Plant transformation vectors are well known in the art. For example, vectors for *Agrobacterium*-mediated transformation are available from the Centre for the Application of Molecular Biology to International Agriculture, GPO Box 3200, Canberra, ACT 2601, Australia (cambia@cambia.org.au).

The present invention also relates to a host cell transformed with a polynucleotide vector construct of the present invention. The host cell can be either prokaryotic or eukaryotic. Bacterial cells are preferred prokaryotic host cells and typically are a strain of *E. coli* such as, for example, the *E. coli* strains DH5 available from Bethesda Research Laboratories Inc., Bethesda, MD, USA, and RR1 available from the American Type Culture Collection (ATCC) of Rockville, MD, USA (No ATCC 31343). Preferred eukaryotic host cells include plant, yeast, insect and mammalian cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human fibroblastic and kidney cell lines. Yeast host cells include YPH499, YPH500 and YPH501 which are generally available from Stratagene Cloning Systems, La Jolla, CA 92037, USA. Preferred mammalian host cells include Chinese hamster ovary (CHO) cells available from the ATCC as CCL61, NIH Swiss mouse embryo cells NIH/3T3 available from the ATCC as CRL 1658, monkey kidney-derived

COS-1 cells available from the ATCC as CRL 1650 and 293 cells which are human embryonic kidney cells. Preferred insect cells are Sf9 cells which can be transfected with baculovirus expression vectors.

Protoplasts for transformation are typically generated as required by methods known in the art. Plant cell lines are not generally available. However, one cell line which is commonly used is the Bright Yellow 2 cell line from tobacco (BY2; Mu *et al* (1997) *Plant Mol. Biol.* **34**, 357-362).

Transformation of appropriate cell hosts with a DNA construct of the present invention is accomplished by well known methods that typically depend on the type of vector used. With regard to transformation of prokaryotic host cells, see, for example, Cohen *et al* (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2110 and Sambrook *et al* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. Transformation of yeast cells is described in Sherman *et al* (1986) *Methods In Yeast Genetics, A Laboratory Manual*, Cold Spring Harbor, NY. The method of Beggs (1978) *Nature* **275**, 104-109 is also useful. With regard to vertebrate cells, reagents useful in transfecting such cells, for example calcium phosphate and DEAE-dextran or liposome formulations, are available from Stratagene Cloning Systems, or Life Technologies Inc., Gaithersburg, MD 20877, USA. With regard to plant cells and whole plants three plant transformation approaches are typically used (J. Draper and R. Scott in D. Grierson (ed.), "Plant Genetic Engineering", Blackie, Glasgow and London, 1991, vol. 1, pp 38-81):

i) *Agrobacterium*-mediated transformation, using the Ti plasmid of *A.*

*tumefaciens* and the Ri plasmid of *A. rhizogenes* (P. Armitage, R. Walden and J. Draper in J. Draper, R. Scott, P. Armitage and R. Walden (eds.), "Plant Genetic Transformation and Expression - A Laboratory Manual", Blackwell Scientific Publications, Oxford, 1988, pp 1-67; R.J. Draper, R. Scott and J. Hamill *ibid.*, pp 69-160);

*Agrobacterium*-mediated transformation is also described in Hooykaas & Schilperoot (1992) *Plant Mol. Biol.* **19**, 15-38; Zupan & Zambryski (1995) *Plant Physiol.* **107**, 1041-1047; and Baron & Zambryski (1996) *Curr. Biol.* **6**, 1567-1569.

ii) DNA-mediated gene transfer, by polyethylene glycol-stimulated DNA uptake into protoplasts, by electroporation, or by microinjection of protoplasts or plant cells (J. Draper, R. Scott, A. Kumar and G. Dury, *ibid.*, pp 161-198). Direct gene transfer into protoplasts is also described in Neuhaus & Spangenberg (1990) *Physiol. Plant* **79**, 213-217; Gad *et al* (1990) *Physiol. Plant* **79**, 177-183; and Mathur & Koncz (1998) *Method Mol. Biol.* **82**, 267-276;

iii) transformation using particle bombardment (D. McCabe and P. Christou, *Plant Cell Tiss. Org. Cult.*, **3**, 227-236 (1993); P. Christou, *Plant J.*, **3**, 275-281 (1992)).

Some species are amenable to direct transformation, avoiding a requirement for tissue or cell culture (Bechtold *et al* (1993) *Life Sciences*, C.R. Acad. Sci. Paris **316**, 1194-1199).

*Agrobacterium*-mediated transformation is generally less effective for

monocotyledonous plants for which approaches ii) and iii) are therefore preferred. However, *Agrobacterium* is capable of transferring DNA to some monocotyledenous plants if tissues containing "competent" cells are infected (see Hiei *et al* (1997) *Plant Mol. Biol.* **35**, 205-218). In all approaches a suitable selection marker, such as kanamycin- or herbicide-resistance, is preferred or alternatively a screenable marker ("reporter") gene, such as β-glucuronidase or luciferase (see J. Draper and R. Scott in D. Grierson (ed.), "Plant Genetic Engineering", Blackie, Glasgow and London, 1991, vol. 1 pp 38-81).

Electroporation is also useful for transforming and/or transfecting cells and is well known in the art for transforming yeast cell, bacterial cells, insect cells, vertebrate cells and some plant cells (eg barley cells, see Lazzeri (1995) *Methods Mol. Biol.* **49**, 95-106).

For example, many bacterial species may be transformed by the methods described in Luchansky *et al* (1988) *Mol. Microbiol.* **2**, 637-646 incorporated herein by reference. The greatest number of transformants is consistently recovered following electroporation of the DNA-cell mixture suspended in 2.5X PEB using 6250V per cm at 25μFD.

Methods for transformation of yeast by electroporation are disclosed in Becker & Guarente (1990) *Methods Enzymol.* **194**, 182.

Successfully transformed cells, ie cells that contain a DNA construct of the present invention, can be identified by well known techniques. For example, cells resulting from the introduction of an expression construct of the present invention can be grown to produce the polypeptide of the

invention. Cells can be harvested and lysed and their DNA content examined for the presence of the DNA using a method such as that described by Southern (1975) *J. Mol. Biol.* **98**, 503 or Berent *et al* (1985) *Biotech.* **3**, 208. Alternatively, the presence of the protein in the supernatant can be detected using antibodies as described below.

In addition to directly assaying for the presence of recombinant DNA, successful transformation can be confirmed by well known immunological methods when the recombinant DNA is capable of directing the expression of the polypeptide. For example, cells successfully transformed with an expression vector produce polypeptides displaying appropriate antigenicity. Samples of cells suspected of being transformed are harvested and assayed for the protein using suitable antibodies.

Thus, in addition to the transformed host cells themselves, the present invention also contemplates a culture of those cells, preferably a monoclonal (clonally homogeneous) culture, or a culture derived from a monoclonal culture, in a nutrient medium.

In relation to plants, it is envisaged that the invention includes single cell derived cell suspension cultures, isolated protoplasts or stable transformed plants. In the latter case it is preferred if the polypeptide of the invention is expressed using an inducible promoter system to avoid potentially lethal effects of gene down-regulation during regeneration of homozygous plants.

Although the polypeptides or polynucleotides of the invention may be introduced into any suitable host cell, it will be appreciated that they are

primarily designed to be effective in appropriate animal or plant cells, particularly those that have one or more sites within their DNA to which the polypeptide of the invention may bind.

Thus, the animal of plant cells which contain a polypeptide (or polynucleotide) of the invention whose presence suppresses the expression of a particular gene, or the animals or plants containing these cells, may be considered to have the gene "knocked out" in the sense that it can no longer be expressed. The chromatin inactivation by histone deacetylation may be irreversible.

It will be readily appreciated that introduction of a polypeptide of the invention into an animal or plant cell, or introduction and expression of a polynucleotide encoding a polypeptide of the invention in an animal or plant cell, will allow targeting of the polypeptide to an appropriate binding site within the DNA (and which is bound by the DNA-binding portion of the polypeptide) and allow for the chromatin at or associated with the target binding site to be inactivated so as to lead to suppression or inactivation of gene expression. Typically, the polypeptide of the invention is selected so that it targets a selected gene. Thus, suitably, the targeted gene has a site which is bound by the DNA binding portion of the polypeptide associated with it. The site which is so bound may be within the gene itself, for example within an intron or within an exon of the gene; or it may be in a region 5′ of the transcribed portion of the gene, for example within or adjacent to a promoter or enhancer region; or it may be in a region 3′ of the transcribed portion of the gene.

Genes regulated by oestrogen receptor (ER) include the progesterone

receptor (PR) gene and the PS2 (trefoil related protein) gene. Thus, the method of the invention may be used to inactivate the PR gene or the PS2 gene when the DNA binding portion of the compound of the invention is at least the DNA-binding portion of ER. Anti-oestrogen therapy is used in the treatment of breast cancer. The full repertoire of oestrogen regulated genes involved in breast cancer is presently unknown. It is generally considered that anti-oestrogen therapy results in the altered expression of key oestrogen regulated genes involved in breast cancer cell growth and transformation. The methods of the invention described below may provide an alternative, potentially more effective, way of regulating the expression (particularly inhibiting) of oestrogen-responsive genes. It may be that for certain DNA binding portions, in a given plant or animal cell there is only one target site and the expression of only one gene is suppressed by the chromatin inactivation portion. However, there may be more than one target site and introduction of a polypeptide (or polynucleotide) of the invention may lead to suppression of expression of a number of genes.

Thus, a further aspect of the invention provides a method of suppressing the expression of a selected gene in a eukaryotic cell the method comprising introducing into the cell (a) a polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with the selected gene which site is present in a eukaryotic genome and a chromatin inactivation portion, or (b) a polynucleotide encoding said polypeptide.

Suitably, the polypeptide is a polypeptide of the invention as described above. Also suitably, the polynucleotide is a polynucleotide of the invention as described above. It is preferred if the preferred polypeptides

or polynucleotides of the invention are used in the method. In particular, it is preferred that the chromatin inactivation portion is PLZF or a portion thereof, for example a portion that can facilitate histone deacetylation. Although it may be a site which has been engineered into the cell, it is preferred if the site at or associated with the selected gene is naturally present in the eukaryotic genome. Preferably, the eukaryotic cell is a plant cell or an animal cell.

The ability to suppress the expression of a selected gene is useful in many areas of biology.

Typically, when the gene whose expression is suppressed is in an animal cell, the animal cell is a cell within an animal and the method of the invention is used to suppress the expression of a selected gene in an animal. For the avoidance of doubt, animal in this context includes human. Examples of particular uses in animal cells include allele-specific inactivation of oncogenic proteins such as mutant *Ras* and mutant *Bcl-10*; inhibition of oestrogen receptor regulated gene expression in breast cancer; inhibition of androgen receptor; inhibition of genes of interest for developmental studies; inhibition of genes for developing transgenic models of human diseases; and inhibition of genes involved in tissue modelling, as found in cancer and wound healing.

Also typically, the plant cell is a cell within a plant and the method of the invention is used to suppress the expression of a selected gene in a plant.

In one embodiment, the method of the invention is used to suppress the expression of socially or environmentally unacceptable or undesirable

genes in commercially engineered transgenic plants. Such genes may include, for example, antibiotic or herbicide selectable marker genes. In this embodiment, the gene in the transgenic plant is targeted for silencing.

In a further embodiment of the invention novel plant architecture or floral morphology may be achieved by targeting some known homeotic genes involved in these developmental pathways.

Suitably, the method of the invention is used to suppress or inactivate the expression of a gene whose expression it is desirable to suppress or inactivate. Such genes include oncogenes, viral genes including genes present in proviral genomes and so the method in relation to animals may constitute a method of medical treatment. Oncogenes may be overexpressed in certain cancers and it may be desirable to suppress their expression. Some oncogenes are oncogenic by virtue of having an activating mutation. Using the method of the invention the selective suppression of expression of a mutant oncogene may be achieved using a DNA binding portion that selectively binds to the mutant oncogene sequence and wherein the chromatin inactivation portion inactivates the chromatin in which the oncogene resides or with which it is associated so that expression of the mutant oncogene is suppressed. Suppression of oncogene overexpression or of mutant (especially activated) oncogene expression is generally desirable in treating cancers in which the oncogenes play a role. Mutant oncogenes which may be targeted by the method of the invention include *Ras* and *Bcl-10*. These may be targeted by engineered DNA binding proteins capable of recognising the mutated genes in a sequence specific manner.

The expression of viral genes in an animal or plant cell is generally undesirable since this expression is often associated with pathogenesis. The nucleic acid of certain viruses may be formed into chromatin and the expression of such viral genes may be controlled by modification of this chromatin. For example, retroviral proviruses (ie DNA copies of retroviral RNAs) are often incorporated into animal and plant genomes where they become part of the chromatin, for example, integrated HIV provirus and integrated human papillomavirus. *Gypsy* and *Copia*-like retrotransposons appear to be widely distributed in the plant kingdom. *Copia*-like retrotransposons, or at least their reverse transcriptase domains, appear broadly distributed in higher plants while the *Gypsy*-like elements (which share their organisation with the retroviruses but lack retroviral envelope domains) are less abundant (Suoniemi *et al* (1998) *Plant J.* **13**, 699-705). Integration of viral DNA into the plant genome has been demonstrated for geminiviral DNA into the tobacco nuclear genome (Bejarano *et al* (1996) *Proc. Natl. Acad. Sci. USA* **93**, 759-764). Potential retroviruses have also recently been described in plants (Wright & Voytus (1998) *Genetics* **149**, 703-715). Using the method of the invention the selective suppression of expression of a viral gene may be achieved. Engineered DNA binding proteins, or RNA binding proteins, such as HIV *tat* protein, may be used to target a chromatin inactivation portion and lead to proviral genome inactivation by binding to nascent genomic RNA transcripts, achieving histone deacetylation by proximity.

Certain genetic diseases are caused by dominant mutations, such as achondroplasia. Suppression of expression of the mutant allele may be useful in treating these diseases. Using the method of the invention the selective suppression of expression of the mutant allele may be achieved

using a DNA binding portion that selectively binds to the mutant allele sequence and wherein the chromatin inactivation portion inactivates the chromatin in which the mutant allele resides or with which it is associated so that expression of the mutant allele is suppressed.

These methods of the invention typically and preferably involve the transfer of a polynucleotide encoding said polypeptide into an animal or plant cell.

Gene transfer systems known in the art may be useful in the practice of the methods of the present invention in which the polynucleotide of the invention is introduced into a cell either within or outwith an animal body. Such an introduction of a polynucleotide may be therapeutically useful and constitutes a form of gene therapy. These include viral and nonviral transfer methods. A number of viruses have been used as gene transfer vectors, including papovaviruses, eg SV40 (Madzak *et al* (1992) *J. Gen. Virol.* **73**, 1533-1536), adenovirus (Berkner (1992) *Curr. Top. Microbiol. Immunol.* **158**, 39-61; Berkner *et al* (1988) *BioTechniques* **6**, 616-629; Gorziglia and Kapikian (1992) *J. Virol.* **66**, 4407-4412; Quantin *et al* (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2581-2584; Rosenfeld *et al* (1992) *Cell* **68**, 143-155; Wilkinson *et al* (1992) *Nucleic Acids Res.* **20**, 2233-2239; Stratford-Perricaudet *et al* (1990) *Hum. Gene Ther.* **1**, 241-256), vaccinia virus (Moss (1992) *Curr. Top. Microbiol. Immunol.* **158**, 25-38), adeno-associated virus (Muzyczka (1992) *Curr. Top. Microbiol. Immunol.* **158**, 97-123; Ohi *et al* (1990) *Gene* **89**, 279-282), herpes viruses including HSV and EBV (Margolskee (1992) *Curr. Top. Microbiol. Immunol.* **158**, 67-90; Johnson *et al* (1992) *J. Virol.* **66**, 2952-2965; Fink *et al* (1992) *Hum. Gene Ther.* 3, 11-19; Breakfield and Geller (1987) *Mol. Neurobiol.*

1, 337-371; Freese *et al* (1990) *Biochem. Pharmacol.* **40**, 2189-2199), and retroviruses of avian (Brandyopadhyay and Temin (1984) *Mol. Cell. Biol.* **4**, 749-754; Petropoulos *et al* (1992) *J. Virol.* **66**, 3391-3397), murine (Miller (1992) *Curr. Top. Microbiol. Immunol.* **158**, 1-24; Miller *et al* (1985) *Mol. Cell. Biol.* **5**, 431-437; Sorge *et al* (1984) *Mol. Cell. Biol.* **4**, 1730-1737; Mann and Baltimore (1985) *J. Virol.* **54**, 401-407; Miller *et al* (1988) *J. Virol.* **62**, 4337-4345), and human origin (Shimada *et al* (1991) *J. Clin. Invest.* **88**, 1043-1047; Helseth *et al* (1990) *J. Virol.* **64**, 2416-2420; Page *et al* (1990) *J. Virol.* **64**, 5370-5276; Buchschacher and Panganiban (1992) *J. Virol.* **66**, 2731-2739). To date most human gene therapy protocols have been based on disabled murine retroviruses.

Nonviral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Graham and van der Eb (1973) *Virology* **52**, 456-467; Pellicer *et al* (1980) *Science* **209**, 1414-1422); mechanical techniques, for example microinjection (Anderson *et al* (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5399-5403; Gordon *et al,* 1980; Brinster *et al* (1981) *Cell* **27**, 223-231; Constantini and Lacy (1981) *Nature* **294**, 92-94); membrane fusion-mediated transfer *via* liposomes (Felgner *et al* (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7413-7417; Wang and Huang (1989) *Biochemistry* **28**, 9508-9514; Kaneda *et al* (1989) *J. Biol. Chem.* **264**, 12126-12129; Stewart *et al* (1992) *Hum. Gene Ther.* **3**, 267-275; Nabel *et al,* 1990; Lim *et al* (1992) *Circulation* **83**, 2007-2011); and direct DNA uptake and receptor-mediated DNA transfer (Wolff *et al* (1990) *Science* **247**, 1465-1468; Wu *et al* (1991) *J. Biol. Chem.* **266**, 14338-14342; Zenke *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3655-3659; Wu *et al,* 1989b; Wolff *et al* (1991) *BioTechniques* **11**, 474-485; Wagner *et al,* 1990; Wagner *et al* (1991) *Proc. Natl. Acad. Sci. USA* **88**,

4255-4259; Cotten *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4033-4037; Curiel *et al* (1991a) *Proc. Natl. Acad. Sci. USA* **88**, 8850-8854; Curiel *et al* (1991b) *Hum. Gene Ther.* 3, 147-154). Viral-mediated gene transfer can be combined with direct *in vivo* gene transfer using liposome delivery, allowing one to direct the viral vectors to the tumour cells and not into the surrounding nondividing cells. Alternatively, the retroviral vector producer cell line can be injected into tumours (Culver *et al* (1992) *Science* **256**, 1550-1552). Injection of producer cells would then provide a continuous source of vector particles. This technique has been approved for use in humans with inoperable brain tumours.

Other suitable systems include the retroviral-adenoviral hybrid system described by Feng *et al* (1997) *Nature Biotechnology* **15**, 866-870, or viral systems with targeting ligands such as suitable single chain Fv fragments.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged.

Liposome/DNA complexes have been shown to be capable of mediating direct *in vivo* gene transfer. While in standard liposome preparations the gene transfer process is nonspecific, localized *in vivo* uptake and expression have been reported in tumour deposits, for example, following direct *in situ* administration (Nabel (1992) *Hum. Gene Ther.* 3, 399-410).

Gene transfer techniques which target DNA directly to a target cell or tissue, is preferred. Receptor-mediated gene transfer, for example, is accomplished by the conjugation of DNA (usually in the form of covalently closed supercoiled plasmid) to a protein ligand *via* polylysine. Ligands are chosen on the basis of the presence of the corresponding ligand receptors on the cell surface of the target cell/tissue type. These ligand-DNA conjugates can be injected directly into the blood if desired and are directed to the target tissue where receptor binding and internalization of the DNA-protein complex occurs. To overcome the problem of intracellular destruction of DNA, coinfection with adenovirus can be included to disrupt endosome function.

It may be advantageous if the polypeptide of the invention is expressed in the target cell using an inducible promoter. Examples of suitable inducible promoters include those that can be induced by heat shock, glucocorticoids, oestradiol and metal ions.

Preferably, the method of suppressing the expression of a selected gene is used to suppress expression of a gene in a human cell; in one particularly preferred embodiment the human cell is within a human body.

However, the method of the invention may involve the modification of animal cells (including human cells) outside of the body of an animal (ie an *ex vivo* treatment of the cells) and the so modified cells may be reintroduced into the animal body.

From the foregoing, it will be appreciated that the method of the invention may be useful to suppress the activity of a plurality of selected genes. In particular, the method of the invention may be used to suppress the activity of a group of genes whose expression is controlled, at least to a large extent, by a single transcription factor. For example, the method may be used to suppress oestrogen-regulated genes as is described in more detail in the Examples.

A further aspect of the invention provides use of a polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with a selected gene which site is naturally present in a eukaryotic genome and a chromatin inactivation portion in the manufacture of an agent for suppressing the expression of the selected gene in a eukaryotic cell.

A still further aspect of the invention provides use of a polynucleotide encoding a polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with a selected gene which site is naturally present in a eukaryotic genome and a chromatin inactivation portion in the manufacture of an agent for suppressing the expression of the selected gene in a eukaryotic cell.

It will be appreciated that it is particularly preferred if the polypeptide or polynucleotide is used in the preparation of a medicament for suppressing the expression of a selected gene in an animal. For the avoidance of doubt, by "animal" we include human.

A further aspect of the invention provides a method of treating a patient in need of suppression of the expression of a selected gene, the method

comprising administering to the patient an effective amount of a polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with the selected gene and a chromatin inactivation portion.

A still further aspect of the invention provides a method of treating a patient in need of suppression of the expression of a selected gene, the method comprising administering to the patient an effective amount of a polynucleotide encoding a polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with the selected gene and a chromatin inactivation portion.

It will be appreciated that suppression of the expression of a selected gene is useful where the expression or overexpression of the selected gene is undesirable and contributes to a disease state in the patient. Examples of undesirable expression of a gene include the expression of certain activated oncogenes in cancer.

Suppression of the expression of the ER upregulated genes is desirable in the treatment of breast cancer. Similarly, suppression of the expression of the androgen receptor (AR)-regulated genes is desirable in the treatment of prostate cancer.

Further aspects of the invention provides use of a polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with a selected gene and a chromatin inactivation portion in the manufacture of a medicament for suppressing the expression of a selected gene in a patient in need of such suppression; and the use of a polynucleotide encoding a

polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with a selected gene and a chromatin inactivation portion in the manufacture of a medicament for suppressing the expression of a selected gene in a patient in need of such suppression.

Still further aspects of the invention provides a polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with a selected gene and a chromatin inactivation portion for use in medicine; and a polynucleotide encoding a polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with a selected gene and a chromatin inactivation portion for use in medicine. Thus, the polypeptide or polynucleotide are packaged and presented for use in medicine.

Yet still further aspects of the invention provide a pharmaceutical composition polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with a selected gene and a chromatin inactivation portion and a pharmaceutically acceptable carrier; and a pharmaceutical composition comprising a polynucleotide encoding a polypeptide comprising a nucleic acid binding portion which binds to a site at or associated with a selected gene and a chromatin inactivation portion and a pharmaceutically acceptable carrier.

By "pharmaceutically acceptable" is included that the formulation is sterile and pyrogen free. Suitable pharmaceutical carriers are well known in the art of pharmacy.

The invention will now be described in more detail with reference to the following Figures and Examples wherein:

Figure 1 shows a schematic representation of the HDAC complex. The schematic representation shows components of the HDAC complex, which includes the Nuclear Receptor Interacting Proteins N-Cor (and SMRT), Sin3, SAP18, SAP30 and RbAp48. It is currently believed that the complex contains HDAC1 and 2 (shown here simply as HDAC). The Nuclear Receptors shown here (Retinoid X Receptor (RXR and its partners Retinoic Acid Receptor α (RAR) or Thyroid Hormone Receptor (TR)) serve as examples of Nuclear Receptors. DNA is represented by the coil and histones as spheres. The arrows emanating from HDAC indicate that HDAC deacetylates histones. The arrow originating from the DNA represents a transcription start site. The postulated effect of histone deacetylation is inhibition of transcription.

Figure 2. Schematic representation of the receptors and PLZF. The domain structure of retinoic acid receptor α (RARα) and oestrogen receptor α (ERα) are shown. Regions A-F refer to regions of differing amino-acid sequence homology as first described by Krust *et al* (1986) *EMBO J.* **5**, 891-897, the DNA binding domain (DBD, region C) and the ligand binding domain (LBD, region E) are most well conserved between the receptors. Transcription Activation Functions 1 and 2 (AF-1 and AF-2) refer to the regions of the receptor containing sequences required for transcriptional activation. The LBD also contains sequences which interact with co-repressor proteins. PLZF-RARα shows the fusion protein in acute promyelocytic leukaemia which results from the t(11;17) translocation, fusing the first 455 amino-acids of PLZF with regions B-F

of RARα (described by Chen *et al* (1993) *EMBO J.* **12**, 1161-1167). The PLZF-ER construct which we have made is shown, fusing the first 455 amino-acids of PLZF with ERα sequences (amino-acids 151-595) homologous to the RARα sequences present in PLZF-RARα. Also shown is HEG19 (see Tora *et al* (1989) *EMBO J.* **8**, 1981-1986 which contains amino-acids 180-595 of human ERα. The open circles with the lettering $Zn^{++}$ show the positions of the Kruppel-like $Zinc^{++}$-binding fingers present in PLZF.

Figure 3. Comparison of the *in vitro* DNA binding by wild-type human ERα and PLZF-ER. (A, B) COS-1 cells were transiently transfected with 5 μg of pSG5 (lane 1), human ERα (HEG0, lane 2), HEG19 (lane 3), PLZF-ERα (lane 4) or PLZF (lane 5). (A) Lysates were prepared in HS buffer and gel shifts performed using [$^{32}$P]-labelled ERE, followed by fractionation on a 5% non-denaturing polyacrylamide gel and visualised by autoradiography. The uncomplexed ERE is seen at the bottom of the gel whilst receptor-complexed ERE migrates more slowly. (B) Lysates prepared for (A) were run on a 10% SDS-polyacrylamide gel, proteins were transferred to nitrocellulose and immunoblotting performed using a monoclonal antibody F3, raised against the C-terminal portion of human ERα. The band at 68 kDal seen in each lane is a cross-reacting protein unrelated to the oestrogen receptor (Ali *et al* (1993) *Hybridoma* **12**, 391-405). Molecular mass marker sizes are listed on the right in kDal.

Figure 4A. PLZF-ERα does not activate an oestrogen-responsive reporter gene in the presence or absence of oestrogen. The oestrogen receptor-negative mammalian COS-1 cell line was transiently co-transfected with 2 μg of the reporter gene 17M-ERE-G-CAT and 1 μg of the β-galactosidase

reporter gene pCH110 (from Pharmacia, UK), together with 0.5 μg of the expression plasmid pSG5 (lanes 1-3), HEG0 (lanes 4-6), HEG19 (lanes 7-9), PLZF-ERα (lanes 10-12) or PLZF (lanes 13-15), as indicated, together with the plasmid BSM to a total DNA concentration of 20 μg. 17β-oestradiol (E, 10 nM) or 4-hydroxy-tamoxifen (O, 100 nM) or prepared in ethanol were added where appropriate. Ethanol was added in the no ligand control. CAT activities were assayed after normalization for β-galactosidase activity from the reference plasmid pCH110. The top panel shows the autoradiographic image. The lower panel shows the result of quantification of four representative experiments. The transcriptional activities are depicted as percentage of the activity observed with HEG0 in the presence of 17β-oestradiol (taken as 100%). For these studies 17β-oestradiol, 4-hydroxytamoxifen and acetyl CoA were purchased from Sigma, UK. $^{14}$C-labelled chloramphenicol was purchased from Amersham International, UK. The other reagents are commonly available from a variety of commercial sources. For more details on the procedures see Metzger *et al* (1995) *Mol. Endocrinol.* 9, 579-591.

Figure 4B. PLZF-ERα represses activation of an oestrogen-responsive reporter gene. The ER-positive breast cancer cell line, MCF-7 was transiently co-transfected with 2 μg of the reporter gene 17M-ERE-G-CAT and 5 μg of the β-galactosidase reporter gene pCH110, together with 0.5 μg of the expression plasmid (pSG5 (lanes 1-3), HEG19 (lanes 4-6), PLZF-ERα (lanes 7-9) or PLZF (lanes 10-12), as indicated, together with the plasmid BSM to a total DNA concentration of 20 μg. Ligands were added and samples processed as described for Figure 4A. The transcriptional activities are depicted as percentage of the activity observed with pSG5 in the presence of 17β-oestradiol (lane 2; taken as 100%).

Figure 5 shows the alignment of the amino-acid sequence encoding HDAC1 from man, arabidopsis and yeast. The aligned amino-acid sequences are shown using the one letter code. The asterisks mark the positions of amino-acids which are identical in all three sequences, whilst the dots show positions of amino-acids which are conservative changes. This alignment is simply meant to illustrate that HDAC genes are present all eukaryotes from yeast and plants to man and is not an exhaustive list of all known sequences. Note further that a total of six mammalian HDAC genes have been described (see Grozinger *et al* (1999) *Proc. Natl. Acad. Sci. USA* **96**, 4868-4873).

Figure 6 shows that PLZF-AR represses activation of an androgen-responsive reporter gene. COS-1 cells were co-transfected with 2 μg of the reporter gene pG29G-tk-CAT and 1 μg of the β-galactosidase reporter plasmid pSGΔlacZ together with 2 μg of the expression plasmid pSG5 (lanes 1-2), or 1 μg of pSG5 (lanes 3-4, 9-12), AR (lanes 3-8), PLZF-AR (lanes 5-6, 9-10) or PLZF (lanes 7-8, 11-12), as indicated, in the presence or absence of R1881 (1 nM), prepared in ethanol. R1881 is a potent activator of the androgen receptor, for example see Quarmby *et al* (1990) *Mol. Endocrinol.* **4**, 1399-4107. Ethanol was added to the control samples (-).

**Example 1: Construction and use of PLZF-ERα gene fusion**

Work was carried out to produce an analogue of the PLZF-RARα gene fusion found in acute promyelocytic leukaemia (Guidez *et al* (1998) *Blood*

**91**, 2634-2642), in which the RARα portion was replaced by an equivalent region of the human estrogen receptor (ERα). To do this, a 1392 bp region of PLZF coding region was amplified by PCR from a full length cDNA clone using a generic oligonucleotide primer to 5′ flanking cloning vector sequence (T7 primer;) and a primer complementary to PLZF sequences encompassed by bases 1441–1446 of the sequence of Chen *et al* (1993) *EMBO J.* **12**, 1161-1167, with additional bases added to the 3′ end, so as to introduce an in-frame XhoI restriction enzyme site (Primer PLZF R; ccgctcgagCTGAATGAGCCAGTAAGTGCATTCTC). Similarly, a 1407 bp region of a human ERα cDNA clone (HEG0; Tora *et al* 1989 *EMBO J.* **8**, 1981-1986) was amplified by PCR using primers which introduced an in frame XhoI site into 5′ coding region and a BamHI site into the 3′ untranslated region (Primers ER F1; CCGCTCGAGggccaaattcagataatcgac and ER R1; ccgtgtgggaTccagggagctctca). PLZF and ER PCR products were restricition enzyme digested with EcoRI and XhoI and XhoI and BamHI respectively. The digest products were purified and ligated with pSG5 expression vector DNA (Stratagene) previously digested with the restriction enzymes EcoR1 and BamHI. The ligation product was used to transform *E. coli* bacteria and plasmid DNA prepared from individual clones. Recombinant pSG5 plasmids containing PLZF-ER gene fusion DNA were initially identified by restriction enzyme digestion and were subsequently confirmed by DNA sequence analysis. The resultant cloned PLZF-ER gene encodes the first 455 amino acids of PLZF, fused in frame with amino acids 151-595 of human ERα sequence. This clone was used in experiments to address expression and subsequent inhibition of ER regulated gene activity by the PLZF-ER fusion protein, as shown in Figures 3 and 4.

Figure 3 shows a comparison of the *in vitro* DNA binding by wild-type human ERα and PLZF-ERα. (A, B) COS-1 cells were transiently transfected with 5 μg of pSG5 (lane 1), human ERα (HEG0, lane 2), HEG19 (lane 3), PLZF-ERα (lane 4) or PLZF (lane 5). (A) Lysates were prepared in HS buffer and gel shifts performed using [$^{32}$P]-labelled ERE, followed by fractionation on a 5% non-denaturing polyacrylamide gel and visualised by autoradiography. The uncomplexed ERE is seen at the bottom of the gel whilst receptor-complexed ERE migrates more slowly. (B) Lysates prepared for (A) were run on a 10% SDS-polyacrylamide gel, proteins were transferred to nitrocellulose and immunoblotting performed using a monoclonal antibody F3, raised against the C-terminal portion of human ERα. The band at 68 kDal seen in each lane is a cross-reacting protein unrelated to the oestrogen receptor. Molecular mass marker sizes are listed on the right in kDal.

The ability of PLZF-ERα to activate transcription of oestrogen responsive genes was tested in the oestrogen receptor-negative COS-1 cells. COS-1 cells were transiently transfected with an oestrogen responsive reporter gene 17M-ERE-G-CAT. Activation of gene expression results in synthesis of the bacterial chloramphenicol acetyl transferase (CAT) protein whose enzymatic activity can be assayed *in vitro* using $^{14}$C-labelled chloramphenicol and acetyl CoA. Acetylation of $^{14}$C-chloramphenicol can be visualised using thin layer chromatography and autoradiography. Quantitation is performed using phosphorimager analysis (Bio-Rad, UK) (see Metzger *et al* (1995) *Mol. Endocrinol.* **9**, 579-591 for details of procedures). Normalization for variations in transfection efficiency is carried out by co-transfection of the reporter gene with a β-galactosidase

expression plasmid (pCH110). As well as the two reporter genes the cells were transfected with the expression vector pSG5 (lanes 1-3), HEG0 (lanes 4-6), HEG19 (lanes 7-9), PLZF-ERα (lanes 10-12) or PLZF (lanes 13-15). Ligands (17β-oestradiol (10 nM, lanes 2, 5, 8, 11, 14) or 4-hydroxytamoxifen (100 nM, lanes 3, 6, 9, 12, 15) were added. The ligands were prepared in ethanol so an equal volume of ethanol was added to the no ligand controls (lanes 1, 4, 7, 10, 13). The results show that the CAT reporter gene activity is increased in the presence of E2 when HEG0 or HEG19 are expressed (lanes 3-6 and 7-9, respectively), HEG19 being a less potent activator than the full-length receptor (HEG0), as expected. PLZF has little if any effect on CAT activity (lanes 13-15). PLZF-ERα also does not activate the CAT reporter gene. Indeed some repression of the background CAT activity seen with pSG5 is observed (lanes 10-12, compare with lanes 1-3). Certainly no activation is seen with PLZF-ERα in the presence of 17β-oestradiol. These results indicate that PLZF-ERα is unable to activate expression of oestrogen responsive genes despite having an ability to bind to oestrogen response elements.

In order to investigate whether PLZF-ERα can inhibit transactivation by endogenous ER, we transiently transfected MCF-7 cells. This is a breast cancer-derived cell line that expresses the oestrogen receptor and requires oestrogen for growth. MCF-7 cells were transiently transfected with an oestrogen responsive reporter gene 17M-ERE-G-CAT and pCH110, together with pSG5 (lanes 1-3), HEG19 (lanes 4-6), PLZF-ERα (lanes 7-9) or PLZF (lanes 10-12) as described for COS-1 cells above. Ligands were added as appropriate (see Figure 4B). The results are displayed in the form of a bar chart. The level of transactivation in the presence of E (Figure 4B, lane 2) was taken as 100%. Transfection of PLZF-ERα

reduced reporter gene activity due to the endogenous ER to below that seen in the absence of ligand (Figure 4B, compare lanes 1 and 7). Addition of 17β-oestradiol (E, lane 8) or the partial antagonist 4-hydroxytamoxifen (O, lane 9) did not result in release from inhibition, suggesting that PLZF-ERα inhibits transactivation in a ligand-independent manner.

From Example 1 it can be concluded that the RARα portion of PLZF-RARα can be replaced by another related DNA binding protein activity.

PLZF-ERα protein retains the ability to bind specifically to the oestrogen receptor DNA binding element. PLZF-ERα displays little, if any ability to activate gene expression.

Replacing the RARα DNA binding activity redirects gene inactivation to the binding specificity of the new DNA binding domain.

PLZF-ERα is able to compete with endogenous ERα in a breast cancer cell lines to over-ride oestrogen-activated gene expression.

PLZF-ERα inhibits oestrogen responsive gene expression in an oestrogen-independent manner.

**Example 2: Construction of PLZF-AR gene fusion and its use**

A second analogue of the PLZF-RARα gene fusion was produced in which the RARα portion was replaced by an equivalent region of the human androgen receptor (AR). A 1146 bp region of a human AR cDNA

clone (Tilley *et al* 1989 *Proc. Natl. Acad. Sci. USA* **86**, 327-331) was amplified by PCR using primers which introduced an in frame XhoI site into the 5′ coding region and a BamHI site immediately following the stop codon (Primers AR F1; ggagctcgagggTTGGAGACTGCCAGGGACC and AR R1; gtgaggatccTCACTGGGTGTGGAAATAGATGG). The AR PCR product was restriction enzyme digested with XhoI and BamHI and ligated with XhoI/BamHI digested PLZF-ER to replace the ER portion with AR. The ligation product was used to transform *E. coli* bacteria and plasmid DNA prepared from individual clones. Recombinant pSG5 plasmids containing PLZF-AR gene fusion DNA were initially identified by restriction enzyme digestion and were subsequently confirmed by DNA sequence analysis. The resultant cloned PLZF-AR gene encodes the first 455 amino acids of PLZF, fused in frame with amino acids 537-917 of human AR. Transient transfections in COS-1 cells, followed by immunoblotting of cell extracts with antibodies directed against PLZF or AR were used to confirm expression and expected size.

Figure 6 shows that PLZF-AR represses activation of an androgen-responsive reporter gene. COS-1 cells were co-transfected with 2 μg of the reporter gene pG29G-tk-CAT and 1 μg of the β-galactosidase reporter plasmid pSGΔlacZ together with 2 μg of the expression plasmid pSG5 (lanes 1-2), or 1 μg of pSG5 (lanes 3-4, 9-12), AR (lanes 3-8), PLZF-AR (lanes 5-6, 9-10) or PLZF (lanes 7-8, 11-12), as indicated, in the presence or absence of R1881 (1 nM), prepared in ethanol. R1881 was purchased from DuPont, USA. Ethanol was added to the control samples (-). The reporter gene has been described in Schule *et al* (1988) *Science* **242**, 1418-1420.

From these experiments it can be concluded that the PLZF-androgen receptor fusion inhibits androgen receptor-mediated transcription activation.

**Example 3: Suppression of androgen receptor-mediated transcription in a prostate cancer patient**

A plasmid vector is produced which encodes the PLZF-AR fusion protein as described in Example 2 under the control of the PSA gene promoter which allows for selective expression in prostate tissue. The plasmid is prepared in a sterile and pyrogen-free form and is formulated into liposomes. The plasmid DNA-containing liposomes are administered into the vicinity of the prostate. Plasmid DNA is taken up by the prostate cancer cells and androgen receptor-mediated transcription is suppressed selectively in prostate cells.

**Example 4: Suppression of oestrogen receptor-mediated transcription in a breast cancer patient**

A retroviral vector is produced which encodes the PLZF-ER fusion protein as described in Example 1. The retroviral vector is administered into the site of the breast tumour. Retroviral RNA is taken up by the breast cancer cells and oestrogen receptor-mediated transcription is suppressed selectively in breast cells.